

UNITED STATES PATENT APPLICATION

OF

Nancy MILLER

Jeffrey SAFFER

Randall SCARBERRY

Scott DECKER

and

Sean ZABRISKI

FOR

**METHODS AND APPARATUS FOR DISPLAYING DISPARATE TYPES OF
INFORMATION USING AN INTERACTIVE SURFACE MAP**

LAW OFFICES

FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.

STANFORD RESEARCH PARK
700 HANSEN WAY
PALO ALTO, CALIF. 94304
650-849-6600

05408716 093099
660660 9T280460

RELATED APPLICATIONS

The following identified U.S. patent applications are relied upon in this application:

U.S. Patent Application Ser. No. _____, entitled "METHODS AND APPARATUS FOR EXTRACTING ATTRIBUTES OF GENETIC MATERIAL," filed on the same date herewith by Jeffrey Saffer, et al.;

U.S. Patent Application Ser. No. 08/713,313, entitled "SYSTEM FOR INFORMATION DISCOVERY," filed on September 13, 1996; and

U.S. Patent Application Ser. No. _____, entitled "DATA PROCESSING, ANALYSIS, AND VISUALIZATION SYSTEM FOR USE WITH DISPARATE DATA TYPES," filed on the same date herewith by Jeffrey Saffer, et al.

The disclosures of each of these applications are herein incorporated by reference in their entirety.

BACKGROUND OF THE INVENTION

A. Field of the Invention

This invention relates generally to methods and apparatus for displaying information graphically.

B. Description of the Related Art

A problem today for many practitioners, particularly in the science disciplines, is the scarcity of available time to review the large volumes of information that are

00408716-090099
660660 " 917280460

LAW OFFICES

FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
STANFORD RESEARCH PARK
700 HANSEN WAY
PALO ALTO, CALIF. 94304
650-849-8600

being collected. For example, modern methods in the life and chemical sciences are producing data at an unprecedented pace. This data may include not only text information, but also DNA sequences, protein sequences, numerical data (e.g., from gene chip assays), and categoric data.

5 Given this flood of diverse information, effective and timely use of the results is no longer possible using traditional approaches, such as lists, tables, or even simple graphs. Furthermore, it is clear that more valuable hypotheses can be derived by simultaneous consideration of multiple types of experimental data (e.g., protein sequence in addition to gene expression data), a process that is currently
10 problematic with large amounts of data.

 Others have developed graphical depictions of multivariate data. See e.g., Nielson GM, Hagen H, Müller H, eds., (1997) Scientific Visualization, IEEE Computer Society, Los Alamitos; Becker RA, Cleveland WS (1987) Brushing Scatterplots, Technometrics 29:127-142; Cleveland WS (1993) Visualizing Data, Hobart Press, Summit, NJ; Bertin J (1983) Seminology of Graphics, University of Wisconsin Press, London; Cleveland WS (1993) Visualizing Data, Hobart Press, Summit, NJ. Although these efforts may provide a graphical description of data,
15 they do not provide an integrated, interactive, and intuitive approach that allows a user to explore information to discover knowledge.

20 There exists, therefore, a need for methods and apparatus that address the shortcomings of these graphical interfaces.

65050" 9T 280460

LAW OFFICES

FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
STANFORD RESEARCH PARK
700 HANSEN WAY
PALO ALTO, CALIF. 94304
650-849-6600

SUMMARY OF THE INVENTION

Methods and apparatus consistent with the present invention, as embodied and broadly described herein, use interactive surface maps to display disparate types of information graphically. These methods and apparatus provide a graphical depiction of records and their attributes in a manner that is easy for the human mind to assimilate, highlights the most informative features of the data, and enables unexpected relationships to be found.

Consistent with the invention, a method of interactively displaying records and their associated attributes involves defining a set of graphic images, wherein each graphic image represents a range of values. The method generates a surface map, with records arranged along a first dimension and graphic images (representing attributes associated with the records) arranged along a second dimension. Upon receiving input from a user selecting a record on the surface map, an index is analyzed to determine if the record is shown in another view. If the record is shown in another view, the visual representation of the record in the other view is altered.

Consistent with the invention, a computer-readable medium includes instructions for controlling a computer system to perform a method for interactively displaying records and their associated attributes. The method involves selecting a set of records and their associated attributes, wherein the associated attributes are any combination of numeric, categoric, sequence, and text information. The method converts the attributes into numeric values, and defines a set of graphic images, wherein each graphic image represents a range of numeric values. The method generates a surface map with the set of records arranged along a first

650-849-6600

LAW OFFICES

FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.

STANFORD RESEARCH PARK
700 HANSEN WAY
PALO ALTO, CALIF. 94304
650-849-6600

dimension and graphic images (representing attributes associated with the records) arranged along a second dimension.

BRIEF DESCRIPTION OF THE DRAWINGS

5 The accompanying drawings, which are incorporated in, and constitute a part of, this specification illustrate an embodiment of the invention and, together with the description, serve to explain the advantages and principles of the invention. In the drawings,

10 FIG. 1 is a block diagram of a system in which methods and apparatus consistent with the present invention may be implemented;

 FIG. 2 is a representative user interface screen showing a galaxy view consistent with the invention;

 FIG. 3 is a flow diagram of a method consistent with the invention for displaying information interactively by using a surface map;

15 FIG. 4a is a representative user interface screen showing a surface map consistent with the invention;

 FIG. 4b is an exploded view of a portion of FIG. 4a;

 FIG. 5 is another representative user interface screen showing a surface map consistent with the invention; and

20 FIG. 6 is another representative user interface screen showing a surface map and a galaxy view consistent with the invention.

09408716-093099
66060-97280460

LAW OFFICES

FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
STANFORD RESEARCH PARK
700 HANSEN WAY
PALO ALTO, CALIF. 94304
650-849-6600

DETAILED DESCRIPTION

Reference will now be made in detail to an embodiment of the present invention as illustrated in the accompanying drawings. The same reference numbers may be used throughout the drawings and the following description to refer to the same or like parts.

A. Overview

Methods and apparatus consistent with the invention provide tools that allow a user to display information interactively so that the user can explore the information to discover knowledge. One such tool displays a set of records and their associated attributes in the form of a detailed, resizeable, scrollable two-dimensional surface map. As used herein, the term "record" (or "object") generally refers to an individual element of a data set. The characteristics associated with records are generally referred to herein as attributes.

The tool also generates reduced-size two- and three- dimensional surface maps that provide an overview of the information displayed in the detailed surface map. Each of these maps are linked to other views, such that a record selected in one map is highlighted in the other views, and vice versa.

B. Architecture

FIG. 1 is a block diagram of a computer system 100 in which methods and apparatus consistent with the invention can be implemented. System 100 comprises a computer 110 connected to a server 180 via a network 170. Network 170 can be, for example, a local area network (LAN), a wide area network (WAN), or the Internet. System 100 is suitable for use with the Java™ programming language, although one skilled in the art will recognize that methods and apparatus

09406716 " 093060
660660 " 9730460

LAW OFFICES

FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.

STANFORD RESEARCH PARK
700 HANSEN WAY
PALO ALTO, CALIF. 94304
650-849-6600

consistent with the invention can be applied to other suitable user environments.

Computer 110 comprises several components that are all interconnected via a system bus 120. Bus 120 can be, for example, a bi-directional system bus that connects the components of computer 110, and contains thirty-two address lines for addressing a memory 125 and a thirty-two bit data bus for transferring data among the components. Alternatively, multiplex data/address lines can be used instead of separate data and address lines. Computer 110 communicates with other users' computers on network 170 via a network interface 145, examples of which include Ethernet or dial-up telephone connections.

Computer 110 contains a processor 115 connected to a memory 125. Processor 115 can be a microprocessor manufactured by Motorola, such as the 680X0 processor, a processor manufactured by Intel, such as the 80X86 or Pentium processors, or a SPARC™ microprocessor from Sun Microsystems, Inc. However, any other suitable microprocessor or micro-, mini-, or mainframe computer, can be used. Memory 125 can include a RAM, a ROM, a video memory, or mass storage. The mass storage can include both fixed and removable media (e.g., magnetic, optical, or magnetic optical storage systems or other available mass storage technology). Memory 125 can include a program, an application programming interface (API), and a virtual machine (VM) that contains instructions for handling constraints, consistent with the invention.

A user typically provides information to computer 110 via a keyboard 130 and a pointing device 135, although other input devices can be used. In return, information is conveyed to the user via display screen 140.

LAW OFFICES

FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.

STANFORD RESEARCH PARK
700 HANSEN WAY
PALO ALTO, CALIF. 94304
650-849-6600

C. Architectural Operation

Before information may be displayed interactively so that a user can explore and discover knowledge, it must be processed into a condition suitable for display. Although this processing is described in detail in U.S. Patent Application Ser. No. _____, entitled "DATA PROCESSING, ANALYSIS, AND VISUALIZATION SYSTEM FOR USE WITH DISPARATE DATA TYPES," it may be described briefly as follows. First, the information represented by the records (including text, numeric, categoric, and sequence / string data) is received in electronic form. Second, the records are analyzed to produce high-dimensional vectors, which are indexed. Third, the high-dimensional vectors are grouped in space to identify relationships. Fourth, the high-dimensional vectors are converted to a two-dimensional representation for viewing purposes, generally referred to herein as "projection." Fifth, the projections may be viewed in different formats according to user-selected options. Each view is linked to an index (or indices), such that a user selection in one view propogates to other views.

One basic visual tool consistent with the invention for viewing information is a "galaxy view," an example of which is shown in Fig. 2. The galaxy view is a two-dimensional scatter graph in which records are organized and depicted in groups (or "clusters") based on relationships between one record and another. In addition to this galaxy view tool, the invention provides numerous interactive visual tools that allow a user to explore and discover knowledge.

Fig. 3 describes one method of displaying information interactively, in the form of a two-dimensional surface map. The method begins with the user selecting a set of records and a set of attributes associated with those records (step 305).

The attributes may comprise any of numerous data types, including the following: numeric, text, sequence (e.g., protein or DNA sequences), or categoric. The selected attributes are converted into numerical values, as explained in U.S. Patent Application Ser. No. _____, entitled "DATA PROCESSING, ANALYSIS, AND VISUALIZATION SYSTEM FOR USE WITH DISPARATE DATA TYPES" (step 310). A set of graphic images are defined, wherein each graphic image represents a range of values (step 315). At one extreme, this range of values may consist of a single value. In one implementation, gray-scale or color rectangular blocks are used as graphic images, with each shade or color representing a distinct range of values. The user may select from a list of predefined color schemes or may independently define a color scheme and its associated range of values.

Next, a two-dimensional surface map is generated to visually depict the records and their associated attributes (step 320). Fig. 4a illustrates one implementation of a resizable, scrollable surface map 405 (the portion of Fig. 4a bounded by "A" and "B") that is arranged as an array, with records forming the rows and attributes forming the columns. Each row within 405, a set of which are shown as 410, depicts information associated with a record. Within each row, a series of gray-scale rectangular blocks are used to depict the value of each attribute associated with that record, as shown in 415.

Fig. 4b is an exploded view of a portion of surface map 405, such as the portion identified as 410 in Fig. 4a. As shown in Fig. 4b, each record is represented by a series of graphic images (such as graphic image 450), that collectively form a row. Each graphic image 450 represents the numeric value of an attribute associated with a record. In short, each "row" of the surface map represents a

record, and each "column" represents the value of a particular attribute for each record.

The ordering of records within map 405 may be defined by the user; or it may be achieved by using algorithms, such as statistical correlation. Similarly, the ordering of the attributes associated with each record may be defined by the user or by an algorithm. Furthermore, relationships between records may be depicted within map 405 in numerous ways. For example, graphical bands (e.g., the two bands shown as 420), may be used to represent related groups of records. Alternatively, conventional dendograms may be used to show relationships between records.

In one implementation, the ordering of records is performed by grouping the records into clusters that have centroids. These clusters are then ordered based on a correlation algorithm applied to the centroids. Finally, within each cluster, the records are ordered by sorting based on the mean distance between each record and the centroids neighboring that record's centroid—the goal being to place each record closest to the neighboring centroid to which it is the most similar. For the terminal clusters, where there is only a single neighboring centroid, the records are sorted by mean distance from the single centroid neighbor. This approach minimizes distances between like records, provides a smooth blending from one record to the next, and allows the user to see structure in the data that would otherwise be difficult to find.

Fig. 4a also shows a reduced-size, two-dimensional surface map 440 (the portion bounded by "C" and "D") that depicts all records and attributes that are

being evaluated. The portion of map 440 that is currently being viewed in enlarged size (i.e., portion 405), is highlighted in 440, as shown by 445. As a result, the reduced-size map 440 provides an overview of the information and allows the user to quickly determine which portion of the information is being shown by map 405.

5 In addition to map 440 shown in Fig. 4a, a three-dimensional surface map 505 may be used, as shown in Fig. 5. In the implementation shown, records are arranged in rows from the bottom-left to the upper-left; attributes are arranged as columns of gray-scale rectangular blocks from the bottom-left to the bottom-right; and values corresponding to each particular attribute for each particular record are represented both by the shade of gray and the height of each rectangular block. Map 505 may contain either the records shown in 405 or all records being evaluated, and may be rotated in any of the three dimensions and/or zoomed to view the information contained therein.

15 In addition to viewing the information in graphical form, the user can interact with the surface maps. Systems consistent with the invention are capable of receiving input from a user selecting a portion of the surface map (step 325). This may be achieved, for example, by using a device to point to a portion of map 405 or by clicking a pointing device on a portion of map 405. In response to this user input, the information associated with the identified portion can be displayed in text format. For example, the record being pointed to in Fig. 4a is identified as "1377T", as shown by 425. Similarly, the attribute being pointed to in Fig. 4a is identified as "META", as shown by 430. The value of the attribute being pointed to in Fig. 4a is identified as "0.0", as shown by 435.

Furthermore, any selections made by the user on a surface map are

09408716-03099
660660-9T280460

LAW OFFICES

FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
STANFORD RESEARCH PARK
700 HANSEN WAY
PALO ALTO, CALIF. 94304
650-849-6600

propagated to other views. For example, in response to receiving input from a user selecting a record in surface map 405, an index is analyzed to determine if the record is shown in another view (step 330). This index is described more fully above in U.S. Patent Application Ser. No. _____, entitled "DATA PROCESSING, ANALYSIS, AND VISUALIZATION SYSTEM FOR USE WITH DISPARATE DATA TYPES." If the record is shown in another display (step 335), the visual representation of that record in the other view is altered (step 340). Fig. 6 is a diagram showing both map 405 and a galaxy view of records 605. If a record is selected on map 405, the record is highlighted in galaxy view 605, and vice versa. Similarly, selecting a group of records on map 405 (as shown by 610) causes the corresponding group of records to be highlighted in galaxy view 605 (as shown by 615), and vice versa.

D. Conclusion

As described in detail above, methods and apparatus consistent with the invention provide tools that allow a user to display information interactively so that the user can explore the information to discover knowledge. The foregoing description of an implementation of the invention has been presented for purposes of illustration and description. Modifications and variations are possible in light of the above teachings or may be acquired from practicing the invention.

For example, although the foregoing description focuses on data types such as text, numeric, categoric, and sequence, those skilled in the art will recognize that other data types may be used consistent with the invention. Furthermore, the foregoing description is based on a client-server architecture, but those skilled in the art will recognize that a peer-to-peer architecture may be used consistent with the

5

invention. Moreover, although the described implementation includes software, the invention may be implemented as a combination of hardware and software or in hardware alone. Additionally, although aspects of the present invention are described as being stored in memory, one skilled in the art will appreciate that these aspects can also be stored on other types of computer-readable media, such as secondary storage devices, like hard disks, floppy disks, or CD-ROM; a carrier wave from the Internet; or other forms of RAM or ROM. The scope of the invention is therefore defined by the claims and their equivalents.

09403716-093099

LAW OFFICES

FINNEGAN, HENDERSON,
FARABOW, GARRETT
& DUNNER, L.L.P.
STANFORD RESEARCH PARK
700 HANSEN WAY
PALO ALTO, CALIF. 94304
650-849-6600